

УДК 517.534

*В роботі розглядається проблема розробки процедури попереднього опрацювання інформації в умовах гетероскедастичності. Пропонується використовувати статистику рекурсивних залишків, щоб підвищити потужність стандартних тестів на перевірку ARCH процесів*

*Ключові слова: гетероскедастичність, волатильність, ризик, GARCH модель, часовий ряд*

*В работе рассматривается проблема разработки процедуры предварительной обработки информации в условиях гетероскедастичности. Предлагается использовать статистику рекурсивных остатков, чтобы повысить мощность стандартных тестов на проверку ARCH процессов*

*Ключевые слова: гетероскедастичность, волатильность, риск, GARCH модель, временной ряд*

*This article represents a problem of developing a procedure for pre-processing information in heteroscedasticity. It is proposed to use the statistics of recursive residuals to improve the capacity of standard tests to check ARCH processes*

*Key words: heteroscedasticity, volatility, risk, GARCH model, time series*

# ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ПОСТРОЕНИЯ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ НЕЛИНЕЙНЫХ ВРЕМЕННЫХ РЯДОВ В УСЛОВИЯХ ГЕТЕРО- СКЕДАСТИЧНОСТИ

**Б. В. Шамша**

Кандидат технических наук, профессор\*  
Контактный тел.: (057) 343-27-78

**А. Ю. Гуд**

Аспирант\*  
Контактный тел.: (097) 746-33-60  
E-mail: work\_pz@i.ua

**А. Н. Одейчук**

Аспирант\*  
Контактный тел.: (057) 335-32-32  
E-mail: anodeychuk@kipt.kharkov.ua  
\*Кафедра информационных управляющих систем  
Харьковский национальный университет  
радиоэлектроники  
пр. Ленина, 14, г. Харьков, Украина, 61166

## 1. Введение. Постановка проблемы в общем виде

Временные ряды финансовых и технико-экономических объектов характеризуются наличием стохастической волатильности, изменчивостью статистических характеристик, в частности дисперсии, которые меняются в зависимости от типа рынка, времени, неконтролируемых возмущений, «тяжелых хвостов» в распределениях, глубины погружения, кластеризации экстремумов, фрактальности, хаотичности и других факторов. В некоторые периоды наблюдается изменчивость в широком и в узком смысле стационарности характеристик. Так, в случае стационарности, имеет место низкая дисперсия, а в другие периоды наблюда-

ется увеличение дисперсии, которая образует своего рода «пучки», «clustering volatility» (кластеризация волатильности).

В таких условиях состояние объекта является неопределенным, рискованным и может привести к тому, что принятие решения в таких условиях становится неэффективным. Функционирование объекта приводит к потерям, убыткам.

Для вычисления меры риска используются различные методы. Заметим, что наиболее распространенными из них являются метод постоянных ковариаций, метод экспоненциально-взвешенных ковариаций, методы, в основе которых лежит построение GARCH моделей, полупараметрических моделей, историче-

ское моделирование, модели, использующие теорию экстремальных значений.

В качестве предметной области в работе предлагается рассмотреть динамические стохастические временные ряды курса валют, которые являются основой для принятия решений в дилинговых информационных системах.

Сложность прогнозирования динамических рядов, характеризующих сущность изменения валютных курсов, биржевых цен, ценных бумаг и других экономических показателей, состоит во временном изменении данных. Действия различных возмущений на такие ряды непостоянны и могут приводить ряд из однородного в неоднородный, и если имеет место коррелированность между наблюдениями в момент времени  $t_1$ , то в момент времени  $t_2$  она может отсутствовать.

Таким образом, с течением времени такой ряд может менять как качественные характеристики, так и количественные. В соответствии с вышеупомянутым, выбор метода прогнозирования будет зависеть от статистических характеристик исходного ряда. Пользователь, на основании информации об эффективности алгоритмов и методов прогнозирования, для каждой конкретной ситуации должен автоматически или автоматизировано выбрать наиболее оптимальный метод.

В этой связи возникают несколько глобальных проблем.

Первая проблема – получение, предварительная обработка исходных данных, разбиение их на классы, текущая оценка состояния временного ряда и отнесение его к тому или другому классу.

Вторая проблема – оценка эффективности методов для каждого состояния данных.

Третья проблема – выбор метода прогноза для конкретного состояния.

Четвертая проблема – оценка адекватности модели прогноза и определение целесообразности использования ее для принятия решений.

Вышеперечисленные проблемы являются основой для создания информационной технологии построения моделей прогнозирования нелинейных процессов и являются актуальными.

## 2. Анализ последних исследований и публикаций

Прогнозирование технико-экономической информации по временным рядам в условиях гетероскедастичности возникает в технике, экономике, социологии, физике, медицине и других предметных областях, и, в частности, при решении задачи оптимизации таких характеристик трафика компьютерной сети, как потери, задержки, джиттер пакетов и т.д. Разработке математических моделей прогнозирования в условиях гетероскедастичности посвящено большое количество работ. Наиболее распространенными являются методы, построенные на базе вероятностно-статистического аппарата. Но их использование требует значительного количества экспериментальных данных, которые не всегда удается собрать. В последнее время в задачах прогнозирования возродился интерес к использованию искусственных нейронных

сетей, важнейшим достоинством которых является возможность обучения и адаптации. Но, как и в случае вероятностно-статистических методов, для обучения нейронных сетей требуется большая выборка экспериментальных данных. Кроме того, обученная нейронная сеть не допускает ясной интерпретации статистических закономерностей, происходящих во временном ряду. Развиваются подходы к прогнозированию, объединяющие экспериментальные данные с экспертно-лингвистической информацией, которая формализуется при помощи нечеткой логики. Использование экспертно-лингвистических закономерностей позволяет увидеть закономерности во временном ряду и построить модель прогнозирования в условиях малых экспериментальных выборок. Основным недостатком таких систем – это невозможность реализовать обучение и адаптацию. Объединение нейронных сетей и принципов управления с нечеткой логикой в настоящее время стали предметом интенсивных исследований, особенно в области прогнозирования. Системы прогнозирования, построенные таким образом, позволяют принимать решения на основе принципов нечеткого мышления с возможностью адаптации и интерпретации результатов. Исходя из вышеизложенного, можно сделать вывод о наличии множества методов прогнозирования, выбор которых является достаточно сложным и наукоемким процессом и предполагает обстоятельное эконометрическое обоснование. Построение модели прогнозирования является многоэтапной процедурой. В рамках ее решения предполагается проведение итерационной процедуры, направленной на решение задач сбора данных и оценки их количества, предварительной статистической обработки, оценки статистических характеристик данных, проверки статистических гипотез, определения структуры временного ряда, выбора эффективного метода прогнозирования, проведения идентификации модели прогнозирования и оценка ее адекватности.

## 3. Цель статьи

В работе основное внимание уделяется разработке процедуры предварительной обработки информации в условиях гетероскедастичности. Предлагается использовать статистику рекурсивных остатков, чтобы повысить мощность стандартных тестов на проверку ARCH процессов.

## 4. Изложение основного материала исследования

При выборе метода прогнозирования приходится решать вопрос об отнесении каждого из рассматриваемых временных рядов к классу стационарных относительно детерминированного тренда (или просто стационарных) – TS (trend stationary) ряды, или к классу рядов, имеющих стохастический тренд (возможно, наряду с детерминированным трендом) и приводящихся к стационарному (или стационарному относительно детерминированного тренда) ряду только путем однократного или  $k$ -кратного дифференцирования ряда – DS (difference stationary) ряды. Для различения

класса стационарности ряда существуют различные критерии, в т.ч. критерии Дики-Фуллера, Филлипса-Перрона, Лейбуна, Квятковского-Филлипса-Шмидта-Шина (или KPSS) [1,2,3].

Под критерием Дики-Фуллера в действительности понимается группа критериев, объединенных одной идеей. В критериях Дики-Фуллера проверяемой (нулевой) является гипотеза о том, что исследуемый ряд принадлежит классу DS (DS-гипотеза); альтернативная гипотеза – исследуемый ряд принадлежит классу TS (TS-гипотеза).

Неправильный выбор оцениваемой статистической модели может существенно отразиться на мощности критерия Дики-Фуллера. Например, если наблюдаемый ряд порождается моделью случайного блуждания со сносом, а статистические выводы делаются по результатам оценивания статистической модели без включения в ее правую часть трендовой составляющей, то тогда мощность критерия, основанная на статистике  $t_\phi$ , стремится к нулю с возрастанием количества наблюдений. С другой стороны, оцениваемая статистическая модель не должна быть избыточной, поскольку это также ведет к уменьшению мощности критерия.

Описанный выше критерий Дики-Фуллера фактически предполагает, что наблюдаемый ряд описывается моделью авторегрессии первого порядка (возможно, с поправкой на линейный тренд). Если же наблюдаемый ряд описывается моделью более высокого (но конечного) порядка  $p$  и характеристический многочлен имеет *не более одного* единичного корня, то тогда можно воспользоваться расширенным (augmented) критерием Дики-Фуллера.

Заметим, что расширенный критерий Дики-Фуллера может применяться и тогда, когда ряд описывается смешанной моделью авторегрессии – скользящего среднего.

Критерий Филлипса-Перрона сводит проверку гипотезы о принадлежности ряда классу DS к проверке гипотезы  $H_0: \phi = 0$  в рамках статистической модели SM:  $\Delta x_t = \phi x_{t-1} + \alpha + \beta t + u_t, t = 2, \dots, T$  где, как и в критерии Дики-Фуллера, параметры  $\alpha$  и  $\beta$  могут быть взяты равными нулю. Однако, в отличие от критерия Дики-Фуллера, случайные составляющие  $u_t$  с нулевыми математическими ожиданиями могут быть автокоррелированными (с достаточно быстрым убыванием автокорреляционной функции), иметь различные дисперсии (гетероскедастичность) и не обязательно нормальные распределения (но такие, что  $E|u_t|^\delta \leq C < \infty$  для некоторого  $\delta > 2$ ). Тем самым, в отличие от критерия Дики-Фуллера, к рассмотрению допускается более широкий класс временных рядов.

Лейбурн предложил вычислять значения статистики критерия Дики-Фуллера DF для исходного ряда и для ряда, получаемого из исходного обращением времени, и затем брать максимум  $DF_{\max}$  из двух полученных значений. Критерий Лейбуна обладает несколько большей мощностью по сравнению с критерием Дики-Фуллера.

Критерий Квятковского-Филлипса-Шмидта-Шина (KPSS) в качестве нулевой берет гипотезу TS. Рассмотрение ведется в рамках модели: Ряд = Детерминированный тренд + Стохастический тренд + Стационарная ошибка.

Стохастический тренд представляется случайным блужданием, и нулевая гипотеза предполагает, что дисперсия инноваций, порождающих это случайное блуждание, равна нулю. Альтернативная гипотеза соответствует предположению о том, что эта дисперсия отлична от нуля, так что анализируемый ряд принадлежит классу DS рядов. В такой формулировке предложенный критерий является LM критерием для проверки указанной нулевой гипотезы. Как и в критерии Филлипса-Перрона, требования на ошибки здесь менее строгие, чем в критерии Дики-Фуллера. Однако при применении данного критерия возникает проблема выбора ширины окна  $l$  в оценке Newey-West, поскольку значения статистики критерия довольно чувствительны к значению  $l$ .

Следующим этапом предварительной обработки является выявление гетероскедастичности. Заметим, что при наличии гетероскедастичности оценки параметров моделей прогнозирования, определяемые классическим методом наименьших квадратов и его модификациями, становятся не эффективными. Для больших выборок статистические оценки не являются оптимальными. В качестве критериев выявления гетероскедастичности используются следующие процедуры: White, Goldfeld-Quandt, Breusch-Pagan, Gltsjer и другие.

Однако, проверка гипотез на наличие гетероскедастичности при использовании вышеуказанных критериев усложняется при наличии непостоянства условного среднего. Возникает проблема формализации проверки на ARCH, когда условное среднее меняется со временем. Решение данной задачи лежит в необходимости разработке теста, робастного к изменчивости условного среднего с использованием информации, доступной для оценки условного среднего.

Предположим, что данные  $\{y_t : t=1, \dots, T\}$  описываются выражением  $y_t = m_t + \varepsilon_t$ , где  $m_t = E(y_t | I_t^*)$  – какая-то непостоянная функция стационарных исходных данных, а  $\varepsilon_t$  имеет нулевое среднее. Условная дисперсия  $h_t = \tilde{\alpha}_0 + \tilde{\alpha}_1 \varepsilon_{t-1}^2 + \dots + \tilde{\alpha}_p \varepsilon_{t-p}^2$  и безусловная дисперсия  $E(\varepsilon_t^2) = \sigma^2$ .

Непостоянство модели с линейной динамикой может быть обусловлено пропущенными сдвигами в функции тренда, выборе длины лага, невозможностью учета нестабильности параметров, автокорреляцией остатков, пропущенными переменными временного ряда. Примечательно, что такая неустановленность в динамических моделях часто приводит к серийным корреляциям в  $\eta_t = m_t - \mu_t$ . Сумма двух стационарных процессов будет серийно коррелирована, если хотя бы в одной серии будет серийная корреляция.

Проблема с непостоянным условным средним представляется как  $\hat{u}_t = \varepsilon_t^2 + \xi_t$ , и очевидно, что наличие  $\varepsilon_t$  приводит к неверной интерпретации наличия ARCH. Тогда понятно, что нашей целью должно быть получение как можно более точной аппроксимации  $m_t$  (в смысле  $L^2$ ).

Пусть условное среднее дается линейной регрессионной моделью  $y_t = z_t' \gamma + u_t$ .

Первое рассматриваемое нами приближение – это включить дополнительные лаги в оцененную модель. Второе приближение необходимо для аппроксимации рекурсивных остатков, так как ненаблюдаемая нелинейность будет всегда проявлена в рекурсивных

остатках. Среди многих стандартных тестов непостоянства модели использование рекурсивных остатков (вместо стандартных OLS остатков) увеличивает мощность тестов.

Для получения рекурсивных остатков необходимо начать с  $k+1$  наблюдения для некоторого предопределенного  $k$  и выполнить рекурсивное оценивание  $u_t$  и  $z_t$  для оставшихся  $T-k$  наблюдений. Процедура тестирования такова, что вначале оценивается AR(1). При оценке AR(1) с постоянным авторегрессионным коэффициентом, тест остатков на ARCH(1) отклоняет нулевую гипотезу отсутствия условной гетероскедастичности, если значение статистики велико. Включение рекурсивных остатков и их квадратов приводит к уменьшению статистики, что уже не является статистически значимым; добавление третьей и четвертой степеней еще уменьшает статистику. В качестве альтернативы, добавление общей суммы (cusum) рекурсивных остатков еще уменьшает статистику.

Присутствует ли ARCH в данных? На самом деле, DGP (data generating process, процесс порождения данных) этих данных, является AR(1) с авторегрессионным коэффициентом 0.3, но содержит небольшое изменение в середине выборки, которое нельзя сразу заметить. Поэтому обнаружение ARCH в этом случае ложное; из-за невозможности учесть измененное среднее мы отвергаем стандартный ARCH тест. Оптимальной процедурой в этом примере является обязательное удаление излома. Действительно, тест на структурное изменение среднего отвергает нулевую гипотезу отсутствия излома. Если мы включим индикатор-переменную в регрессию, чтобы учесть небольшое изменение в реальной дате излома, значение теста статистики на ARCH будет повышаться. Если имеет место излом, необходимо проверить и удалить его, прежде чем провести проверку на ARCH. Однако, необнаруженный излом в целом может и не нанести ущерб, если рассматривать робастные процедуры.

Одним из подходов для получения остатков, которые бы несли достоверную информацию о наличии гетероскедастичности, является применение нейронных сетей. Заметим, что нейронные сети достаточно хорошо аппроксимируют нелинейные функции.

Остатки, полученные с использованием нейронных сетей, представляют собой набор данных, обеспечивающих применение стандартных критериев обнаружения гетероскедастичности.

---

## 5. Выводы

---

Предлагается использовать статистику рекурсивных остатков, чтобы повысить мощность стандартных тестов на проверку ARCH. Моделирование методом Monte Carlo показало, что использование рекурсивных остатков позволило удалить искажения. Более того, удалось сравнить мощности критериев обнаружения гетероскедастичности при различных уровнях нелинейности, объемах выборок и остатков, полученных методом МНК, ВМНК, максимального правдоподобия, наименьших модулей, нейронных сетей. Результаты моделирования показали, что выбор критерия обнаружения гетероскедастичности существенно зависит от величины нелинейности. При изменении характера нелинейности большинство критериев меняют свои оценки.

---

## Литература

---

1. Engle, R.F. Estimates of the variance of US Inflation Based on the ARCH model // Journal of Money, Credit and Banking, v.15, 1983, pp.286-301.
2. Bollerslev, Tim, Ray At. Chou and Kenneth F. Kroner. ARCH Modeling in Finance: And review of the theory and empirical evidence // Journal of econometrics, v.52, 1992, pp.5-59.
3. Шамша Б.В., Гуржий А.М., Дудар З.В., Левикін В.М. Математичне забезпечення інформаційно-управляючих систем // Підручник для студентів ВНЗ. – Харьков: ТОВ Компанія СМІТ, 2005. – 448 с.